PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
H04B 3/23

A1

(11) International Publication Number: WO 99/41848

(43) International Publication Date: 19 August 1999 (19.08.99)

(21) International Application Number: PCT/SE99/00135

(22) International Filing Date: 2 February 1999 (02.02.99)

(30) Priority Data:
09/023,304 13 February 1998 (13.02.98) US

(71) Applicant: TELEFONAKTIEBOLAGET LM ERICSSON (publ) [SE/SE]; S–126 25 Stockholm (SE).

(72) Inventors: TRUMP, Tõnu; Bäverbäcksgränd 95, S–124 62 Stockholm (SE). ERIKSSON, Anders; Smålandsvägen 2, S–757 58 Uppsala (SE).

(74) Agent: ERICSSON RADIO SYSTEMS AB; Common Patent Dept., S–164 80 Stockholm (SE).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

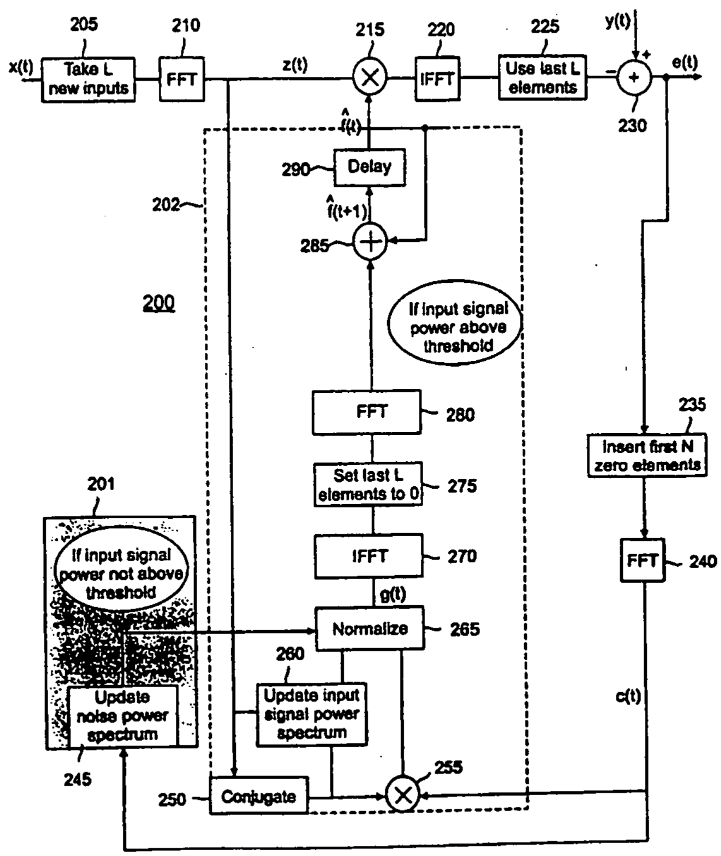
**Published**

*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: METHODS AND APPARATUS FOR CONTROLLING FILTER ADAPTION IN NOISE

(57) Abstract

Methods and apparatus for incorporating partial knowledge of system measurement noise into an adaptive filtering algorithm, for example in the context of an echo cancelation system. Adaptive algorithms according to the invention are obtained by minimizing a best linear unbiased estimate (BLUE) criterion function using a stochastic gradient method and by then converting to the frequency domain to reduce computational complexity. Advantageously, system noise characteristics are measured during natural pauses in user speech and then taken into account during filter adaptation. As a result, echo cancelers constructed according to the invention can provide good echo cancelation even in situations where there is considerable background noise. According to an exemplary embodiment, an echo cancelation device configured to suppress an echo component of an observed signal, wherein the echo component results from coupling of an echo source signal through an echo path. The echo cancelation device includes an adaptive filter configured to approximate the echo path and to thereby provide an estimate of the echo component, wherein an adaptive algorithm of the adaptive filter incorporates a measurement of a noise component of the observed signal.



*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | | | TR | Turkey |
| BG | Bulgaria | HU | Hungary | ML | Mali | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MN | Mongolia | UA | Ukraine |
| BR | Brazil | IL | Israel | MR | Mauritania | UG | Uganda |
| BY | Belarus | IS | Iceland | MW | Malawi | US | United States of America |
| CA | Canada | IT | Italy | MX | Mexico | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NE | Niger | VN | Viet Nam |
| CG | Congo | KE | Kenya | NL | Netherlands | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NO | Norway | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | NZ | New Zealand | | |
| CM | Cameroon | | | PL | Poland | | |
| CN | China | KR | Republic of Korea | PT | Portugal | | |
| CU | Cuba | KZ | Kazakstan | RO | Romania | | |
| CZ | Czech Republic | LC | Saint Lucia | RU | Russian Federation | | |
| DE | Germany | LI | Liechtenstein | SD | Sudan | | |
| DK | Denmark | LK | Sri Lanka | SE | Sweden | | |
| EE | Estonia | LR | Liberia | SG | Singapore | | |

# METHODS AND APPARATUS FOR CONTROLLING FILTER ADAPTION IN NOISE

## BACKGROUND OF THE INVENTION

The present invention relates to communications systems, and more particularly, to echo suppression in bi-directional communications links.

Signal echo is a well known phenomenon by which energy in a transmitted signal reflects back toward the source of the signal. For example, in the context of network telephony, signal echo can result from reflections at the hybrid circuits which link four-wire network connections to two-wire local connections (e.g., at a subscriber loop). Considerable echo is also generated when a telephone set is used in the so-called hands-free mode (e.g., in teleconferencing systems and automobile telephone applications). In the hands-free context, echo results from acoustical coupling between a telephone loudspeaker and a proximate telephone microphone.

Such signal echo is not typically a concern where the round trip delay associated with the echo is small. However, where the round trip delay is longer (e.g., on the order of hundreds of milliseconds), the signal echo can become annoying or even intolerable for a human telephone user. In practice, the round trip delay can be a physical delay caused by a long transmission path (e.g., when a satellite link is used) or a processing delay caused by network equipment (e.g., a digital speech encoder or decoder).

The echo problem can be understood with reference to Figure 1, wherein a bi-directional communication link between a far-end user and a near-end user is depicted in combination with an exemplary echo suppression system 100 in which the teachings of the present invention can be implemented. As shown, the exemplary echo suppression system 100 includes an echo canceler 110 comprising an adaptive filter 105, a double-talk detector 115, a summation device 125 and a non-linear processor 135.

In Figure 1, a far-end input signal $x(t)$ (e.g., an audio signal in the context of network telephony) is coupled to a reference input of the adaptive filter 105 and to a first reference input of the double-talk detector 115. Additionally, an echo estimate output $\hat{s}(t)$ of the adaptive filter 105 is coupled to a subtractive input of the summation device 125, and an observed near-end signal $y(t)$ is coupled to an additive input of the

detection output of the double-talk detector 115 is coupled to a control input of the adaptive filter 105, and an error signal output $e(t)$ of the summation device 125 is coupled to an input of the non-linear processor 135. An output of the non-linear processor 135 serves as a far-end signal output for the echo canceler 110.

In operation, the far-end signal $x(t)$ is provided to the near-end user and is at the same time modified and reflected back toward the far-end user by the near-end echo path H. As described above, the near-end echo path H results, for example, from a hybrid mismatch in a telephone network and/or an unblocked acoustic path between a near-end loudspeaker and a near-end microphone. The near-end echo path H can be characterized by its impulse response **h** and is represented conceptually as an finite impulse response (FIR) filter 150 in Figure 1.

As shown, the resulting echo signal $s(t)$ then combines with a near-end input signal $v(t)$ to yield the observed near-end signal $y(t)$ at the input of the echo canceler 110. The near-end signal $v(t)$ includes, for example, near-end voice and/or near-end background noise. Mixing of the near-end signal $v(t)$ and the echo signal $s(t)$ is represented in the system 100 of Figure 1 by way of a second summing device 155. Those skilled in the art will appreciate, however, that the second summing device 155 is conceptual in nature and that no such device is actually included in a practical system. Rather, mixing of the near-end signal $v(t)$ and the echo signal $s(t)$ is inherent in the system configuration (e.g., when a near-end microphone simultaneously picks up near-end voice and noise as well as echo from a near-end loudspeaker).

To prevent the echo signal $s(t)$ from reaching the far-end user, the echo canceler 110 dynamically adjusts the impulse response of the adaptive filter 105 to match that of the near-end echo path H and subtracts the resulting echo estimate $\hat{s}(t)$ from the observed near-end signal $y(t)$ to provide the echo-canceled error, or residual signal $e(t)$. However, since the adaptive filter 105 typically cannot provide a perfect match (e.g., due to non-linearities in the network equipment and/or dynamic changes in the near-end echo path), the echo canceler also uses the non-linear processor, or residual echo suppressor 135 to provide additional echo suppression as necessary. Additionally, because the signals involved in a telephone call are nonstationary by nature, the echo canceler 110 also typically uses the double-talk detector 115 to inhibit filter adaptation when a

measurement of the echo signal to near end signal ratio indicates that improvement of the echo path estimate cannot be attained by continuing to adapt the filter 105.

According to most conventional systems, the echo canceler 110 employs a least mean squares (LMS) adaptive algorithm or a normalized least mean squares (NLMS) adaptive algorithm to update the filter 105. Such algorithms are attractive for use in practical systems as they are quite robust and provide low computational complexity. Other conventional systems have employed the well known affine projection adaptive algorithm (see, for example, U.S. Patent No. 5,428,562 to Gay, issued June 27, 1995) and various frequency domain algorithms (see, for example, J.J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *IEEE Signal Processing Magazine*, Jan. 1992, pp. 14-37 and J. Soo and K.K. Pang, "Multidelay Block Frequency Domain Adaptive Filter," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 38, No. 2, Feb. 1990, pp. 373-376).

Most of the above described conventional adaptive algorithms are designed to minimize in some sense the error signal power $E[e^2(t)]$, or some estimate thereof, and therefor implicitly assume that the noise component of the near-end input signal $v(t)$ is either absent or white. While such an assumption is accurate where the noise component consists primarily of thermal noise or quantization noise, it is often faulty in practical situations where the noise component instead consists primarily of background sounds from the near-end environment.

Consequently, conventional echo canceling systems often perform poorly in the presence of substantial colored near-end noise, and thus there is a need for improved methods and apparatus for adapting echo cancelation filters.

## SUMMARY OF THE INVENTION

The present invention fulfills the above-described and other needs by providing methods and apparatus for incorporating partial knowledge of system measurement noise into an adaptive filtering algorithm. Exemplary algorithms are obtained by minimizing a best linear unbiased estimate (BLUE) criterion function using a stochastic gradient method and by then converting to the frequency domain to reduce computational complexity. Advantageously, system noise characteristics are measured during natural

pauses in user speech and then taken into account during filter adaptation. Thus, as compared to conventional devices, an echo canceler constructed according to the invention provides superior echo cancelation in situations where there is considerable background noise.

According to an exemplary embodiment, an echo cancelation device is configured to suppress an echo component of an observed signal, wherein the echo component results from coupling of an echo source signal through an echo path. According to the embodiment, the echo cancelation device includes an adaptive filter configured to approximate the echo path and to thereby provide an estimate of the echo component, wherein an adaptive algorithm of the adaptive filter incorporates a measurement of a noise component of the observed signal. Advantageously, the measurement of the noise component of the observed signal can be an *a priori* measurement made in an environment in which the echo cancelation device is designed to operate. Alternatively, measurements of the noise component of the observed signal can be made in real time during the natural pauses in one or more speech components of the observed signal. In an exemplary embodiment, such pauses are determined based upon a measurement of power in a block of samples of the echo source signal.

According to additional embodiments, the echo estimate is subtracted from the observed signal to provide an echo-canceled error signal, and an estimated impulse response of the adaptive filter is computed based upon an estimate of a power spectrum of the error signal and an estimate of a power spectrum of the echo source signal. Advantageously, the estimated impulse response of the adaptive filter can be selectively updated. For example, the estimated impulse response can be updated when a measurement of power in a block of samples of the echo source signal is above a threshold, and not updated otherwise. According to embodiments, the estimated impulse response of the adaptive filter is computed based upon a weighted sum of the error signal power spectrum estimate and the echo source power spectrum estimate. For example, the weighted sum can be computed by adding the echo source power spectrum estimate to a product of the error signal power spectrum estimate and a noise matching parameter. Advantageously, the noise matching parameter can be adjusted dynamically based upon samples of the error signal and samples of the echo source signal.

An exemplary method for canceling an echo component of an observed signal according to the invention includes the steps of sampling an echo source signal to provide a block of source signal samples, computing a discrete Fourier transform of the block of source signal samples to provide a frequency domain representation of the source signal, multiplying the frequency domain representation of the source signal by a frequency domain estimate of an impulse response of the echo path to provide a frequency domain echo estimate, computing an inverse discrete Fourier transform of the frequency domain echo estimate to provide a time domain echo estimate, subtracting the time domain echo estimate from the observed signal to provide an echo canceled error signal, computing a discrete Fourier transform of a block of samples of the error signal to provide a frequency domain representation of the error signal, and adapting the frequency domain estimate of the impulse response of the echo path based upon the frequency domain representation of the error signal and based upon the frequency domain representation of the source signal. Advantageously, the step of adapting the frequency domain estimate of the impulse response incorporates a measurement of a noise component of the observed signal.

According to exemplary embodiments, the step of adapting the frequency domain estimate of the impulse response includes the steps of updating the frequency domain estimate of the impulse response when a measurement of power in the block of source signal samples is above a threshold, and updating an estimate of a noise power spectrum of the error signal otherwise. Additionally, the step of updating the frequency domain estimate of the impulse response can include the steps of updating an estimate of a signal power spectrum of the source signal, computing a weighted sum of the noise power spectrum estimate and the signal power spectrum estimate, and computing an update of the frequency domain estimate based upon the weighted sum. Advantageously, the weighted sum can be computed by adding the signal power spectrum estimate to a product of the noise power spectrum estimate and a noise matching parameter. According to embodiments, the noise matching parameter is dynamically adjusted based upon samples of the echo source signal and the echo canceled error signal.

These and other features and advantages of the present invention are explained in greater detail hereinafter with reference to the illustrative examples shown in the accompanying drawings. Those skilled in the art will appreciate that the described

embodiments are provided for purposes of illustration and understanding and that numerous equivalent embodiments are contemplated herein.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts an exemplary echo cancelation system in which embodiments of the invention can be implemented.

Figure 2 depicts an exemplary adaptive algorithm according to the invention.

Figures 3-5 depict computer simulation results demonstrating the efficacy of embodiments of the invention.

Figures 6 and 7 depict an alternative adaptive algorithm according to the invention.

## DETAILED DESCRIPTION OF THE EXEMPLARY EMBODIMENTS

According to the invention, a best linear unbiased estimate (BLUE) criterion function is used as a target function in the development of adaptive filter algorithms for use in echo cancelation systems. Advantageously, such a target function enables the resulting algorithms to account for color which can be present in the measurement noise of practical echo cancelation systems. Consequently, the resulting algorithms provide improved echo cancelation performance as compared to conventional algorithms.

In the following rigorous development of the adaptive algorithms of the present invention, italic letters, boldface lowercase letters and boldface uppercase letters are used to represent scalars, column vectors and matrices, respectively. Additionally, superscripts $T$, $H$ and $\dagger$ are used to denote transpose, Hermitian transpose and Moore-Penrose pseudo-inverse operations, respectively. Further, the symbol **I** is used to represent the identity matrix, and the symbol **0** is used to represent a matrix of all zeros.

Assuming that the echo path H in an echo canceling system can be modeled as a finite impulse response filter having N taps or filter coefficients, and also assuming that the correlation properties of the near-end input signal $x(t)$ of the system can be well modeled using L correlation lags, then an alternative best linear unbiased estimate criterion function $V$ can be defined as:

$$V = (\mathbf{Xh} - \mathbf{y})^T(\gamma\mathbf{R}_v + \mathbf{XX}^T)^\dagger(\mathbf{Xh} - \mathbf{y}), \tag{1}$$

where $\mathbf{h}$ is the $N$ vector of unknown filter coefficients (i.e., the impulse response of the echo path H), $\gamma$ is a positive constant and $\mathbf{X}$ is an $M \times N$ matrix of input signal samples defined as $\mathbf{X} = [\mathbf{x}(t)...\mathbf{x}(t - N + 1)]$, with $\mathbf{x}(t) = [x_{t-M+1}, ... , x_t]^T$. Additionally, the vector $\mathbf{y} = \mathbf{Xh} + \mathbf{v}$ is the $M$ vector of observed signal samples, and $\mathbf{v}$ is the $M$ vector of measurement noise samples having symmetric Toeplitz covariance matrix $\mathbf{R}_v$. Note that, although the above provided definition of the input sample matrix $\mathbf{X}$ is perhaps not the most consequent definition in certain contexts, it beneficially provides frequency domain algorithms which show strong similarities with the classical overlap-save frequency domain adaptive filter (FDAF) algorithms which are described, for example, in G.A. Clark, S.R. Parker and S.K. Mitra, "A Unified Approach to Time- and Frequency-Domain Realization of FIR Adaptive Digital Filters," *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol. 31, pp. 1073-1083, Oct. 1983; S. Haykin, "*Adaptive Filter Theory*," Third Edition, Prentice Hall, 1996; and the above cited article by J.J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *IEEE Signal Processing Magazine*, Jan. 1992, pp. 14-37).

Given the criterion function definition of equation (1), the derivative of the criterion function $V$ with respect to the unknown coefficient vector $\mathbf{h}$ can be written as

$$\begin{aligned}\frac{\partial V}{\partial \mathbf{h}} &= 2[\mathbf{X}^T(\gamma\mathbf{R}_v + \mathbf{XX}^T)^\dagger\mathbf{Xh} - 2\mathbf{X}^T(\gamma\mathbf{R}_v + \mathbf{XX}^T)^\dagger\mathbf{y} \\ &= -2\mathbf{X}^T(\gamma\mathbf{R}_v + \mathbf{XX}^T)^\dagger\mathbf{e},\end{aligned} \tag{2}$$

where $\mathbf{e} = \mathbf{y} - \mathbf{Xh}$, corresponding to samples of the error signal $e(t)$ described above. An estimate of the coefficient vector $\mathbf{h}$ can thus be written as

$$\hat{\mathbf{h}} = (\mathbf{X}^T(\gamma\mathbf{R}_v + \mathbf{XX}^T)^\dagger\mathbf{X})^{-1}\mathbf{X}^T(\gamma\mathbf{R}_v + \mathbf{XX}^T)^\dagger\mathbf{y}. \tag{3}$$

Advantageously, it can be shown that equation (3) is the best linear unbiased estimate of the coefficient vector $\mathbf{h}$ for the case of possibly singular $\mathbf{R}_v$. Further, it is argued in T. Söderström and P. Stoica, "*System Identification*," Prentice Hall, 1988, pp. 89-90, that this estimate, at least for the case $\gamma = 1$, has better numerical properties than the more traditional best linear unbiased estimate resulting from minimization of

$(\mathbf{Xh} - \mathbf{y})^T\mathbf{R}_v^{-1}(\mathbf{Xh} - \mathbf{y})$ (see, for example, K.C. Ho, "A Minimum Misadjustment Adaptive FIR Filter," *IEEE Trans. Signal Processing*, Vol. 44, pp. 577-585, March, 1996).

According to the invention, a gradient algorithm for estimating the coefficient vector $\mathbf{h}$ can then be defined such that equation (1) is minimized. Specifically, a gradient algorithm according to the invention is defined as

$$\mathbf{e}(t) = \mathbf{y}(t) - \mathbf{X}(t)\hat{\mathbf{h}}(t) \tag{4}$$

$$\hat{\mathbf{h}}(t+1) = \hat{\mathbf{h}}(t) + \mu\mathbf{X}^T(t)(\gamma\mathbf{R}_v+\mathbf{X}(t)\mathbf{X}^T(t))^{\dagger}\mathbf{e}(t). \tag{5}$$

Note that in the equations above, the length $M$ of the entire signal has been replaced by a block length $L < M$ in all corresponding matrix dimensions. Note also that, in the case of white noise $v(t)$, the gradient algorithm of the invention coincides with the relaxed and regularized form of the affine projection algorithm with the regularization parameter proportional to the measurement noise variance (see, for example, S.L. Gay and S. Tavathia, "The Fast Affine Projection Algorithm," *Proc. ICASSP '95*, Detroit, Vol. 5, pp. 3023-3026, May, 1995).

According to the invention, however, the equations above are used to develop a frequency domain algorithm. To do so, an $(N + L) \times (N + L)$ cyclically extended input signal matrix is defined as

$$\mathbf{X}_c(t) = \mathrm{cycl}([x_{t-N-L+1}\; x_t\; x_{t-1} \;\ldots\; x_{t-N-L+2}])$$

$$= \begin{bmatrix} x_{t-N-L+1} & x_t & \cdots & x_{t-N-L+2} \\ x_{t-N-L+2} & x_{t-N-L+1} & \cdots & x_{t-N-L+3} \\ & & \vdots & \\ x_{t-L} & x_{t-L+1} & \cdots & x_{t-L-1} \\ x_{t-L+1} & x_{t-L} & \cdots & x_{t-L-2} \\ & & \vdots & \\ x_t & x_{t-1} & \cdots & x_{t-N-L+1} \end{bmatrix}.$$

Note that this is just one particular cyclic extension of several possible. The advantage of this cyclic extension is that it leads to intuitively well understandable algorithms. A similar extension is used, for example, in the above cited paper by G.A.

Clark, S.R. Parker and S.K. Mitra, "A Unified Approach to Time- and Frequency-Domain Realization of FIR Adaptive Digital Filters," *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol. 31, pp. 1073-1083, Oct. 1983.

The original input signal matrix $\mathbf{X}(t)$ appears in the lower left corner of the cyclically extended input signal matrix $\mathbf{X}_c(t)$ and, consequently, the input signal matrix $\mathbf{X}(t)$ can be written as

$$\mathbf{X}(t) = [\mathbf{0}_{L\times N}\,\mathbf{I}_L]\mathbf{X}_c(t)\begin{bmatrix}\mathbf{I}_N\\ \mathbf{0}_{L\times N}\end{bmatrix}. \tag{6}$$

The extended matrix $\mathbf{X}_c(t)$ is right-cyclic by construction and so is its transpose $\mathbf{X}_c^T(t)$. The eigendecomposition of a right-cyclic matrix (see, for example, S.L. Marple, Jr., "*Digital Spectral Analysis with Applications*," Prentice Hall, 1987) is given by

$$\mathbf{X}_c^T = \mathbf{F}\Lambda\mathbf{F}^H, \tag{7}$$

where the symbol $\mathbf{F}$ represents the $(N+L)\times(N+L)$ discrete Fourier transform matrix having elements $F_{kl} = \frac{1}{\sqrt{N+L}}exp\left(\frac{-j2\pi kl}{N+L}\right)$, and where the symbol $\Lambda$ represents a diagonal matrix formed by the discrete Fourier transform of the first row of the extended transpose matrix $\mathbf{X}_c^T$. Specifically, the diagonal matrix $\Lambda$ is defined as

$$\Lambda = \mathrm{diag}(\mathrm{DFT}([x_{t-N-L+1}\ldots x_t])).$$

In the case of real valued signals, the transpose of $\mathbf{X}_c$ and the Hermitian transpose of $\mathbf{X}_c$ are the same and, hence, $\mathbf{X}_c^T = \mathbf{F}\Lambda\mathbf{F}^H = \mathbf{F}^H\Lambda^H\mathbf{F}$ and $\mathbf{X}_c = \mathbf{F}\Lambda^H\mathbf{F}^H = \mathbf{F}^H\Lambda\mathbf{F}$.

Thus, substituting equations (6) and (7) into equations (4) and (5), the adaptive algorithm of the invention can be expressed as:

$$\mathbf{e}(t) = \mathbf{y}(t) - [\mathbf{0}_{L\times N}\,\mathbf{I}_L]\mathbf{F}^H\Lambda\mathbf{F}\begin{bmatrix}\mathbf{I}_N\\ \mathbf{0}_{L\times N}\end{bmatrix}\hat{\mathbf{h}}(t) \tag{8}$$

$$\hat{\mathbf{h}}(t+1) = \hat{\mathbf{h}}(t) + \mu[\mathbf{I}_N \mathbf{0}_{N\times L}]\mathbf{F}^H \Lambda^H \mathbf{F} \tag{9}$$

$$\times \begin{bmatrix} \mathbf{0}_{N\times L} \\ \mathbf{I}_L \end{bmatrix} (\gamma\mathbf{R}_v + \mathbf{X}\mathbf{X}^T)^{\dagger}\mathbf{e}(t).$$

Further, by defining a frequency response vector for the unknown system as

$$\hat{\mathbf{f}}(t) = \mathbf{F}\begin{bmatrix} \mathbf{I}_N \\ \mathbf{0}_{L\times N} \end{bmatrix}\hat{\mathbf{h}}(t), \tag{10}$$

and by defining an intermediate vector $\mathbf{g}(t)$ as

$$\mathbf{g}(t) = \mathbf{F}\begin{bmatrix} \mathbf{0}_{N\times L} \\ \mathbf{I}_L \end{bmatrix} (\gamma\mathbf{R}_v + \mathbf{X}(t)\mathbf{X}^T(t))^{\dagger}\mathbf{e}(t), \tag{11}$$

the adaptive algorithm of the invention can be expressed as

$$\mathbf{e}(t) = \mathbf{y}(t) - [\mathbf{0}_{L\times N}\ \mathbf{I}_L]\mathbf{F}^H\ \Lambda\hat{\mathbf{f}}(t) \tag{12}$$

$$\hat{\mathbf{f}}(t+1) = \hat{\mathbf{f}}(t) + \mu\mathbf{F}\begin{bmatrix} \mathbf{I}_N \\ \mathbf{0}_{L\times N} \end{bmatrix}[\mathbf{I}_N\ \mathbf{0}_{N\times L}]\mathbf{F}^H\ \Lambda^H\mathbf{g}(t). \tag{13}$$

According to a first exemplary embodiment, the term $\gamma\mathbf{R}_v + \mathbf{X}(t)\mathbf{X}^T(t)$ is approximated by a Toeplitz matrix, and a recursive algorithm such as the well known Levinson algorithm (see, for example, the above cited text by S.L. Marple, Jr., "*Digital Spectral Analysis with Applications*," Prentice Hall, 1987) is used to find the term $(\gamma\mathbf{R}_v + \mathbf{X}(t)\mathbf{X}^T(t))^{\dagger}\mathbf{e}(t)$. This first embodiment is referred to hereinafter as ALGO1.

For the case of $L \ll N$, the first embodiment ALGO1 provides satisfactory computational complexity on the order of $(O((N + L)\log_2(N + L)) + O(L^2))$. However, since the complexity of the Levinson algorithm is not acceptable in all practical situations, the present invention utilizes several simplifying approximations to provide an

alternative, more streamlined algorithm. Advantageously, the above described first embodiment ALGO1 has been used in computer simulations to validate the approximations used to derive the lower complexity algorithm. Results of the computer simulations are described in detail below with reference to Figures 3-5.

To derive the lower complexity algorithm, the intermediate vector $\mathbf{g}(t)$ is expressed alternatively as

$$\mathbf{g}(t) = \left(\mathbf{F}\begin{bmatrix}\mathbf{0}_{N\times L}\\ \mathbf{I}_L\end{bmatrix}(\gamma\mathbf{R}_v + \mathbf{X}(t)\mathbf{X}^T(t)) \times [\mathbf{0}_{L\times N}\ \mathbf{I}_L]\ \mathbf{F}^H\right)^{\dagger}\mathbf{F}\begin{bmatrix}\mathbf{0}_{N\times 1}\\ \mathbf{e}(t)\end{bmatrix}. \tag{14}$$

To see that this is possible, consider the equation $\mathbf{Rz} = \mathbf{e}$, where $\mathbf{R} = \gamma\mathbf{R}_v + \mathbf{XX}^T$ is a nonnegative definite $L \times L$ matrix and $\mathbf{z}$ and $\mathbf{e}$ are $L$ vectors. The latter portion of the equation is equivalent to

$$\begin{bmatrix}\mathbf{0}_N & \mathbf{0}_{N\times L}\\ \mathbf{0}_{L\times N} & \mathbf{R}\end{bmatrix}\begin{bmatrix}\mathbf{0}_{N\times 1}\\ \mathbf{z}\end{bmatrix} = \begin{bmatrix}\mathbf{0}_{N\times 1}\\ \mathbf{e}\end{bmatrix},$$

which can be rewritten as

$$\mathbf{F}\begin{bmatrix}\mathbf{0}_N & \mathbf{0}_{N\times L}\\ \mathbf{0}_{L\times N} & \mathbf{R}\end{bmatrix}\mathbf{F}^H\mathbf{F}\begin{bmatrix}\mathbf{0}_{N\times 1}\\ \mathbf{z}\end{bmatrix} = \mathbf{F}\begin{bmatrix}\mathbf{0}_{N\times 1}\\ \mathbf{e}\end{bmatrix}.$$

Thus, by equation (11), the vector $\mathbf{g}(t)$ can be expressed as

$$\mathbf{g}(t) = \mathbf{F}\begin{bmatrix}\mathbf{0}_{N\times 1}\\ \mathbf{z}\end{bmatrix}$$

or

$$\mathbf{g}(t) = \left( \mathbf{F} \begin{bmatrix} \mathbf{0}_N & \mathbf{0}_{N\times L} \\ \mathbf{0}_{L\times N} & \mathbf{R} \end{bmatrix} \mathbf{F}^H \right)^\dagger \mathbf{F} \begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{e} \end{bmatrix},$$

which is equivalent to the expression for $\mathbf{g}(t)$ provided in equation (14).

According to the invention, the input signal correlation matrix is then approximated by a Toeplitz Matrix $\mathbf{R}_x \approx \mathbf{X}(t)\mathbf{X}^T(t)$, and a cyclic extension $\mathbf{R}_c$ of the sum of the correlation matrices is defined to satisfy

$$\gamma\mathbf{R}_v + \mathbf{X}(t)\mathbf{X}^T(t) \approx \gamma\mathbf{R}_v + \mathbf{R}_x = [\mathbf{0}_{L\times N}\, \mathbf{I}_L]\mathbf{R}_c \begin{bmatrix} \mathbf{0}_{N\times L} \\ \mathbf{I}_L \end{bmatrix}.$$

Then, assuming that $L \leq N+1$, the first row of the extension $\mathbf{R}_c$ obeys even DFT symmetry and, as a result, it can be decomposed as

$$\mathbf{R}_c = \mathbf{F}^H\mathbf{D}\mathbf{F} = \mathbf{F}\mathbf{D}\mathbf{F}^H, \tag{15}$$

where $\mathbf{D}$ is a diagonal matrix with the discrete Fourier transformed first row of $\mathbf{R}_c$ on the main diagonal. The main diagonal of $\mathbf{D}$ is thus real and symmetric by construction. Accordingly, equation (14) can now be rewritten as

$$\mathbf{g}(t) = (\mathbf{KDK})^\dagger \mathbf{F} \begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{e}(t) \end{bmatrix}, \tag{16}$$

where

$$\mathbf{K} = \mathbf{F} \begin{bmatrix} \mathbf{0}_{N\times L} \\ \mathbf{I}_L \end{bmatrix} [\mathbf{0}_{L\times N}\mathbf{I}_L]\mathbf{F}^H$$

can be seen as a window matrix. More specifically, the matrix $\mathbf{K}$ is a right-cyclic matrix with elements of the first row given by

$$\mathbf{K}(1,n) = \frac{1}{N+L}\exp\left(\frac{-j\omega(2N+L-1)}{2}\right)\frac{\sin(\omega L/2)}{\sin(\omega/2)},$$

$\omega = \dfrac{2\pi n}{N+L}$, which corresponds to a rectangular window in the time domain.

Note that using equation (16) in equation (13) still results in an algorithm with relatively high computational complexity. Advantageously, however, the present invention provides a very low complexity algorithm by approximating the matrix **K** by an identity matrix. Interestingly, such an approximation results in an algorithm which is very similar to the self-orthogonalizing frequency domain adaptive filter (see, for example, the above cited article by J.J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *IEEE Signal Processing Magazine*, Jan. 1992, pp. 14-37) with signal power at frequency bins (used to normalize the step sizes at respective frequencies) replaced by the weighted sum of signal and measurement noise powers.

In the foregoing discussion, it has been assumed that the measurement noise correlation matrix $\mathbf{R}_v$ is known. In practice, however, $\mathbf{R}_v$ is often unknown and is estimated. For example, in voice echo cancelation applications, it is natural to initialize the estimate as $\sigma_v^2\mathbf{I}$, where $\sigma_v^2$ is an expected noise power. Further, in certain applications, it is possible to obtain a better initial estimate during the algorithm initialization phase. The estimate can then be refreshed during natural pauses in speech (e.g., when $\mathrm{Tr}(\Lambda\Lambda^H)$ is above a predefined threshold *th*). Advantageously, doing so does not have a significant impact on the computational complexity of the algorithm, as the coefficients should not be updated when the input signal power is low anyway (see, for example, T. Petillon, A. Gilloire and S. Theodoridis, "The Fast Newton Transversal Filter: An Efficient Scheme for Acoustic Echo Cancellation in Mobile Radio," *IEEE Trans. Signal Processing*, Vol. 42, pp. 509-517, March 1994).

The resulting low-complexity algorithm according to the invention, referred to hereinafter as ALGO2, is thus described arithmetically as follows:

## ADAPTIVE ALGORITHM ALGO2

INITIALIZE ESTIMATES AS:

$$\hat{\mathbf{f}}(t=0) = \mathbf{0}_{(N+L)\times 1}$$
$$\mathbf{P}_x(t=0) = \sigma_x^2 \mathbf{I}_{N+L}$$
$$\mathbf{P}_v(t=0) = \sigma_v^2 \mathbf{I}_{N+L}$$

THEN, FOR EACH NEW BLOCK

OF $L$ INPUT SAMPLES, COMPUTE:

$$\Lambda = \mathrm{diag}(\mathrm{DFT}([x_{t-N-L+1} \cdots x_t]))$$
$$\mathbf{e}(t) = \mathbf{y}(t) - [\mathbf{0}_{L\times N}\,\mathbf{I}_L]\mathbf{F}^H\ \Lambda\hat{\mathbf{f}}(t)$$
$$\mathbf{c}(t) = \mathbf{F}\begin{bmatrix}\mathbf{0}_{N\times 1}\\ \mathbf{e}(t)\end{bmatrix}$$

AND, IF $\mathrm{Tr}(\Lambda\Lambda^H) > th$, THEN COMPUTE:

$$\mathbf{P}_x = \beta\,\mathbf{P}_x + (1-\beta)\,\Lambda\Lambda^H$$
$$\mathbf{D} = \mathbf{P}_x + \gamma\mathbf{P}_\upsilon$$
$$\mathbf{g}(t) = \mathbf{D}^\dagger\mathbf{c}(t)$$

$$\hat{\mathbf{f}}(t+1) = \hat{\mathbf{f}}(t) + \mu\mathbf{F}\begin{vmatrix}\mathbf{I}_N\\ \mathbf{0}_{L\times N}\end{vmatrix}[\mathbf{I}_N\mathbf{0}_{N\times L}]\mathbf{F}^H\Lambda^H\mathbf{g}(t)$$

ELSE, COMPUTE:

$$\mathbf{P}_\upsilon = \beta\,\mathbf{P}_\upsilon + (1-\beta)\,\mathrm{diag}(\mathrm{c})\mathrm{diag}(\mathrm{c})^H$$

Note that the equations of the low-complexity algorithm ALGO2 can be applied every sample or once per block of samples. Applying the equations once every block of $L$ samples reduces computational complexity at the price of creating an $L$-sample delay and providing a somewhat slower initial convergence. Note also that the matrix **D** can be

updated directly in the frequency domain instead of first updating the matrix $\mathbf{R}_c$ and computing the DFT.

A flow diagram corresponding to the above described low-complexity algorithm ALGO2 is depicted in Figure 2. As shown, the input signal $x(t)$ is coupled to an input of a first sampling block 205, and an output of the first sampling block is coupled to an input of a first FFT block 210. An output $\mathbf{z}(t)$ of the first FFT block 210 is coupled to a first input of a first multiplier 215, and an output of the first multiplier 215 is coupled to an input of a first IFFT block 220. An output of the first IFFT block 220 is coupled to an input of a second sampling block 225, and an output of the second sampling block 225 is coupled to a subtractive input of a first summing device 230. The observed near-end input signal $y(t)$ is coupled to an additive input of the first summing device 230, and an output of the first summing device 230 serves as the residual or error signal $e(t)$.

The error signal $e(t)$ is coupled to an input of a first zero-padding block 235, and an output of the first zero-padding block 235 is coupled to an input of a second FFT block 240. An output $\mathbf{c}(t)$ of the second FFT block 240 is coupled to an input of a noise power spectrum update block 245 and to a first input of a second multiplier 255. The second multiplier 255 receives an additional input from a conjugation block 250, and the conjugation block 250 in turn receives the output $\mathbf{z}(t)$ of the first FFT block 210 as input. The output $\mathbf{z}(t)$ of the first FFT block 210 and an output of the conjugation block 250 are also coupled to first and second inputs, respectively, of a signal power spectrum update block 260. An output of the signal power spectrum update block 260 and an output of the second multiplier 255 are then coupled to first and second inputs, respectively, of a normalization block 265. Additionally, an output of the noise power spectrum update block 245 is coupled to a third input of the normalization block 265.

An output $\mathbf{g}(t)$ of the normalization block 265 is coupled to an input of a second IFFT block 270, and an output of the second IFFT block 270 is coupled to an input of a second zero-padding block 275. An output of the second zero-padding block 275 is coupled to an input of a third FFT block 280, and an output of the third FFT block 280 is coupled to a first additive input of a second summing device 285. An output $\hat{\mathbf{f}}(t+1)$ of the second summing device 285 is coupled to an input of a one-sample delay block 290, and

an output $\hat{\mathbf{f}}(t)$ of the one-sample delay block 290 is coupled to a second input of the first multiplier 215 and to a second additive input of the second summing device 285.

In operation, the first sampling block 205 forms a block of L new and N old samples of the input signal $x(t)$. The first FFT block 210 then computes the discrete Fourier transform of the resulting input sample block (preferably using a Fast Fourier Transform algorithm) to provide the frequency domain sample vector $\mathbf{z}(t)$. The first multiplier 215 then element-wise multiplies the vector $\mathbf{z}(t)$ by the delay block output $\hat{\mathbf{f}}(t)$ (which represents the frequency-domain estimate of the impulse response $\mathbf{h}$) to form an echo estimate in the frequency domain. Thereafter, the first IFFT block 220 computes the inverse discrete Fourier transform of the resulting frequency-domain echo estimate, and the second sampling block 225 extracts the last L samples to provide the time-domain echo estimate (i.e., $\hat{s}(\mathrm{t})$). The first summation device 230 then subtracts the resulting time-domain echo estimate from the observed near-end signal $y(t)$ to form a sample block of the residual signal $\mathbf{e}(t)$ as desired.

As shown, the residual signal $\mathbf{e}(t)$ is used in combination with the frequency-domain sample vector $\mathbf{z}(t)$ to compute the frequency-domain impulse response estimate $\hat{\mathbf{f}}(t)$. Specifically, the first zero-padding block 235 complements the residual signal sample block with N leading zeros, and the second FFT block computes the discrete Fourier transform of the resulting zero-padded sample block to form the frequency domain representation of the residual signal (i.e., $\mathbf{c}(t)$). Thereafter, a determination is made as to whether the input signal power level is large enough to generate a significant amount of echo. Specifically, the short-term power estimate (i.e., based on the current block of samples of the frequency domain input signal $\mathbf{z}(t)$) is compared to the predetermined threshold $th$ as described above. If the input signal power is above the threshold $th$, the algorithm proceeds to update the frequency-domain impulse response estimate as indicated by a lightly shaded box 202 in Figure 2. Otherwise, the algorithm updates the noise power spectrum estimate as indicated by a darkly shaded box 201 in Figure 2. Of course, the adaptation can also be qualified using the output of a double talk detector as described above with reference to Figure 1.

When the impulse response estimate is to be updated, the signal power spectrum update block 260 first computes a new estimate of the input signal power spectrum $\mathbf{p}_x$

(corresponding to the elements on the main diagonal of the signal power matrix $\mathbf{P}_x$ described above). In exemplary embodiments, the signal spectrum estimate is computed by element-wise multiplying the frequency domain input vector $\mathbf{z}(t)$ with its complex conjugate (i.e., the output of the conjugation block 250) and then averaging using, for example, an exponential window. The second multiplier 255 then element-wise multiplies the frequency-domain residual signal $\mathbf{c}(t)$ by the complex conjugate of the vector $\mathbf{z}(t)$, and the normalization block 265 normalizes (i.e., element-wise divides) the result using a weighted sum of the input and noise power spectra (i.e., $\mathbf{p}_x + \gamma\, \mathbf{p}_v$ ) to form the intermediate vector $\mathbf{g}(t)$. Selection of the weighting parameter $\gamma$, referred to hereinafter as the noise matching parameter, is described in detail below.

Once the intermediate vector $\mathbf{g}(t)$ is determined, the second IFFT block 270 computes the inverse discrete Fourier transform of the vector $\mathbf{g}(t)$, and the second zero-padding block 275 sets the last L elements of the resulting vector to zero. Thereafter, the third FFT block 280 computes the discrete Fourier transform of the output of the second zero-padding block 275, and the resulting vector is multiplied by the step size, or update gain $\mu$ and added (at the second summing device 285) to the previous frequency-domain impulse response estimate $\hat{\mathbf{f}}(t+1)$ to provide an updated frequency-domain impulse response estimate $\hat{\mathbf{f}}(t)$ as desired.

When the impulse response estimate is not to be updated (i.e., when the input signal power is not above the threshold *th*, indicating that there is a pause in user speech), the estimate of noise power spectrum $\mathbf{p}_v$ (corresponding to the elements on the main diagonal of the noise power matrix $\mathbf{P}_v$ described above) is updated based on the frequency domain representation of the residual signal $\mathbf{c}(t)$. In exemplary embodiments, the noise spectrum estimate $\mathbf{p}_v$ is computed by element-wise multiplying the frequency domain representation of the residual signal $\mathbf{c}(t)$ with its complex conjugate and then averaging using, for example, an exponential window. By periodically updating the noise power estimate in this way, any color present in the noise signal will be accounted for when the impulse response estimate is adapted.

Thus, the flow diagram of Figure 2 depicts a practical implementation of the low-complexity algorithm ALGO2 of the invention. Those skilled in the art will appreciate that there is a direct correspondence between the blocks of Figure 2 and the elements of

the equations used above to define the exemplary algorithm ALGO2. Those skilled in the art will also appreciate that the operational blocks of Figure 2 can be implemented in practice using, for example, standard digital signal processing devices, application specific integrated circuitry, or a general purpose digital computer.

Computer simulations were performed in order to compare the performance of the low complexity algorithm ALGO2 with that of the first, reference algorithm ALGO1 and with that of the frequency domain adaptive filter described in the above cited article by J.J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *IEEE Signal Processing Magazine*, Jan. 1992, pp. 14-37. The computer simulations also tested the relative performance of all three algorithms with respect to actual speech signals.

First, stationary signals were used to directly evaluate the validity of the approximations made when deriving the streamlined algorithm ALGO2 from the reference algorithm ALGO1. Stationary signals were also used to directly compare the performance of the exemplary algorithms ALGO1, ALGO2 with the performance of the above cited frequency domain adaptive filter. For purposes of these simulations, stationary signals are those for which the signal statistics do not depend on time for $t > 0$.

A typical simulation result is shown in Figure 3, wherein a first plot 310 shows the learning curves (i.e. $E[(e(t) - v(t))^2]$) for all three algorithms, and a second plot 320 shows the corresponding weight errors $E[\frac{\|\mathbf{h}-\hat{\mathbf{h}}(t)\|^2}{\|\mathbf{h}\|^2}]$. The curves represent ensemble averages taken over 200 independent trials. For the simulations shown in Figure 3, the input signal $x(t)$ is an autoregressive process generated by

$$\frac{1}{1+0.7z^{-1}-0.1z^{-2}+0.1z^{-3}+0.01z^{-4}},$$

and the measurement noise is also an autoregressive process generated by

$$\frac{1}{1-0.93z^{-1}+0.01z^{-2}}.$$

Additionally, the SNR is approximately 0dB, and the true impulse response is a 64-tap FIR filter with flat frequency response and coefficients distributed uniformly in the interval $[-1/128, 1/128]$. The design parameters were chosen as follows: $N = 64$, $L = 64$,

$\mu = 0.04$, $\beta = 0.99$, and $\gamma = 128$. Since the simulated signals are stationary, the measurement noise spectrum is estimated first (this time is not shown in Figure 3) and then used in the adaptive algorithm.

In Figure 3, note that the curves corresponding to the first algorithm ALGO1 (i.e., curves 312, 322) are not distinguishable from those corresponding to the low-complexity algorithm ALGO2 (i.e., curves 313, 323). Advantageously, this indicates that the approximations made in developing the low-complexity algorithm ALGO2 are at least reasonably good. Further, note in Figure 3 that both of the exemplary algorithms ALGO1, ALGO2 outperform the frequency domain adaptive filter (curves 311, 321).

Additional computer simulations were conducted to study the performance of the streamlined algorithm ALGO2 in an acoustic echo cancelation application (specifically in an automobile hands-free telephone application). Figure 4 presents the learning curves (averaged in time using an exponential window) of one typical simulation. For these simulations, the design parameters were $N = 256$ (corresponding to a 256-tap echo impulse response identified in a Volvo 940 with two people sitting in the front seats), $L = 256$, $\mu = 0.04$, $\beta = 0.99$, and $\gamma = 128$. The noise power spectra $\mathbf{P}_v$ was estimated on-line. For the observed input signal $y(t)$, a female speaker is generating the echo, and the echo is then corrupted by a noise recorded in a moving automobile. The sampling rate is 8000 Hz, and the curves in Figure 4 thus correspond to a one minute time interval.

As shown in the first plot 410 of Figure 4, the echo power is lower than the noise power most of the time. However, as the signals are concentrated in different frequency bands, the echo is clearly audible before adaptive processing. Though the frequency domain adaptive filter does improve the situation, the echo remains audible even after processing as shown in the second plot 420 of Figure 4. By way of contrast, the low-complexity algorithm of the invention ALGO2 makes the residual echo barely audible in the noise (after an initial convergence period), as shown in the third plot 430 of Figure 4.

Figure 5 shows the power spectra computed using samples 394000 to 410000 from the curves of Figure 4. It can be seen from the first plot 510 that the echo signal power is above the measurement noise power at most frequencies. As shown in the second plot 520, processing by the frequency domain adaptive filter attenuates echo, but the residual echo power remains above the measurement noise power at several

frequencies. By way of contrast, the third plot 530 shows that processing by the low complexity algorithm ALGO2 reduces the residual echo power below the measurement noise power at all frequencies, and thus the masking effects of human perception make the echo inaudible.

Note that the exemplary low-complexity algorithm ALGO2 introduces an L-sample delay into the signal processing path. Since it is usually desirable to choose L approximately equal to the impulse response length N in order to reduce computational complexity, such a delay may be unacceptable for long echo path impulse responses.

Advantageously, however, additional embodiments of the invention reduce the signal processing delay without dramatically increasing computational complexity. Specifically, the alternative embodiments divide the echo path impulse response **h** into sections of length K and treat the adaptive filter as a series connection of the resulting sectional filters. An analogous approach is described, for example, in the above cited paper by J. Soo and K.K. Pang, "Multidelay Block Frequency Domain Adaptive Filter," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 38, No. 2, Feb. 1990, pp. 373-376).

According to the additional embodiments, it is advantageous to set the section length K approximately equal to the block length L. As K<<N, the delay is reduced. In many applications the echo canceler is followed by a speech coder which operates on blocks of signal samples and introduces a delay larger than its block length. The main contribution to this delay comes from collecting the input signal samples. In these applications, it is desirable to set L equal to the speech codec block length, or to an integer fraction thereof. Then the echo canceler output block can be delivered to the speech coder, instead of a sequence of samples, and the overall delay is only slightly increased. The resulting algorithm, referred to hereinafter as ALGO3, is as follows:

## ADAPTIVE ALGORITHM ALGO3

INITIALIZE ESTIMATES AS:

$$\hat{\mathbf{f}}_k(t=0) = \mathbf{0}_{(K+L)\times 1}, \forall k$$

$$\mathbf{P}_x(t=0) = \sigma_x^2 \mathbf{I}_{K+L}$$

$$\mathbf{P}_v(t=0) = \sigma_v^2 \mathbf{I}_{K+L}$$

THEN, FOR EACH NEW BLOCK OF $K$ INPUT SAMPLES, COMPUTE:

$$\Lambda_k = \Lambda_{k+1}, \forall k \neq kmax$$

$$\Lambda_{kmax} = \mathrm{diag}(\mathrm{DFT}([x_{t-K-L+1} \ldots x_t]))$$

$$\mathbf{e}(t) = \mathbf{y}(t) - \sum_k [0_{L\times K} \mathbf{I}_L]\mathbf{F}^H \Lambda_k \hat{\mathbf{f}}_k(t)$$

$$\mathbf{c}(t) = \mathbf{F}\begin{bmatrix} 0_{K\times 1} \\ \mathbf{e}(t) \end{bmatrix}$$

AND, IF $\sum_k \mathrm{Tr}(\Lambda_k \Lambda_k^H) > th$, THEN COMPUTE:

$$\mathbf{P}_x = \beta \mathbf{P}_x + (1-\beta) \sum_k \Lambda_k \Lambda_k^H$$

$$\mathbf{D} = \mathbf{P}_x + \gamma \mathbf{P}_v$$

$$\mathbf{g}(t) = \mathbf{D}^{\dagger} \mathbf{c}(t)$$

$$\hat{\mathbf{f}}_k(t+1) = \hat{\mathbf{f}}_k(t) + \mu \mathbf{F}\begin{bmatrix} I_K \\ 0_{L\times K} \end{bmatrix} [\mathbf{I}_K 0_{K\times L}]\mathbf{F}^H \Lambda_k^H \mathbf{g}(t), \forall k$$

ELSE, COMPUTE:

$$\mathbf{P}_v = \beta \mathbf{P}_v + (1-\beta)\,\mathrm{diag}\,(\mathbf{c})\,\mathrm{diag}(\mathbf{c})^H$$

A flow diagram corresponding to the above described algorithm ALGO3 is depicted in Figures 6 and 7. Those skilled in the art will appreciate that, just as the operational blocks of Figure 2 correspond to the elements of the equations used to define the low-complexity algorithm ALGO2, so do the operational blocks of Figures 6 and 7 correspond to the elements of the equations used to define the alternative algorithm ALGO3. Like the operational blocks of Figure 2, those of Figures 6 and 7 can be

implemented in practice using, for example, standard digital signal processing devices, application specific integrated circuitry, or a general purpose digital computer.

Note that each of the exemplary algorithms ALGO1, ALGO2, ALGO3 includes several design parameters which are set before the algorithm is implemented. Advantageously, the invention provides a guideline for properly setting these design parameters in practice.

For example, the filter length N can be set based on the particular application at hand. In network echo cancelation and acoustic echo cancelation for automobile hands-free applications, a filter length or model order in the range of 256 to 512 taps typically provides quality results. On the other hand, in teleconferencing applications in which the impulse response of an entire conference room is modeled, a filter length on the order of several thousand taps typically provides better results.

The impulse response section length K, used in the alternative algorithm ALGO3, establishes the processing delay of the algorithm and should therefore be made relatively small. On the other hand, the overall computational complexity of the algorithm increases as K decreases, and a proper trade-off is therefore established based on available computing power. In applications in which an echo canceler is followed by a speech coder, it is typically advantageous to set K equal to the block length of the speech coder, or some integer fraction thereof. Additionally, K is typically selected such that the number of filter sections (i.e., N/K) is an integer.

The least computationally intensive implementation results from setting the block length L approximately equal to the filter length N in the first low-complexity embodiment ALGO2 and approximately equal to the section length K in the alternative low-complexity embodiment ALGO3. It is also advantageous to set N+L or K+L equal to an integer power of two in the first and second low-complexity embodiments ALGO2, ALGO3, respectively.

The parameter $\beta$ determines the memory length of the power spectrum estimates and is made positive and less than one. Selection of $\beta$ in the interval [0.9, 0.995] has shown to provide quality results.

The step size $\mu$ is made positive and small enough to guarantee stability of the adaptive algorithm. Making the step size too small, however, slows down the adaptation,

especially in combination with long impulse responses. Setting the step size in the range from $0.1/R$ to $0.5/R$, where $R = \sqrt{N+L}$ for the first low-complexity embodiment ALGO2 and $R = \sqrt{K+L}$ for the alternative low-complexity embodiment ALGO3, has proven to work well.

As noted above, the noise matching parameter $\gamma$ weights the normalizing power spectra of the input signal and the measurement noise and is made nonnegative. Making the noise matching parameter too large will cause the measurement noise to be weighted too heavily and will thus significantly slow down the convergence of the algorithm. On the other hand, making the noise matching parameter too small will cause the measurement noise to be weighted too lightly and will thus provide relatively fast adaptation at the price of higher residual echo power at steady state. In such case, the shape of the residual echo spectrum will be close to the original echo spectrum with some degree of attenuation (i.e., approximately the same echo attenuation will be obtained over the entire frequency band). By way of contrast, reasonably high values of $\gamma$ will lead to improved performance at steady state in the sense that the residual echo spectrum will be further attenuated near the frequencies where it is above the measurement noise spectrum (i.e., the masking effects of human perception tend to render the echo inaudible when the entire residual echo spectrum lies below the noise spectrum). The price for such an improvement can be a slightly slower initial convergence of the algorithm. However, according to the invention, this effect can be overcome by making the parameter $\gamma$ time dependent.

For example, the noise matching parameter $\gamma$ can be computed as

$$\gamma(t) = \frac{E[x^2(t)]}{E[e^2(t)] - E[v^2(t)]},$$

where E[] denotes mathematical expectation. This approach provides a relatively small $\gamma$ when the residual echo is much stronger than the noise and thus provides relatively fast adaptation. However, if the residual echo power approaches the noise power, then $\gamma$

becomes large, and the algorithm starts to color the residual echo spectrum as described above. In practice, the mathematical expectations are not known and are estimated. One example of such an estimate is a sum of squares over a block containing the frequency domain representation of the corresponding signal. In other words, the noise matching parameter can be computed as

$$\gamma(t) = \frac{\sum |z(t)|^2}{\mathrm{Max}\left( \left(\sum |c(t)|^2 - \sum |c_v|^2\right), const\right)},$$

where $c_v$ is a block containing the frequency domain representation of the measurement noise and const is a small positive constant guaranteeing a positive $\gamma$ in the presence of estimation errors.

In sum, the present invention provides a number of adaptive algorithms which can be used, for example, in echo canceling applications. The exemplary algorithms are derived by minimizing a best linear unbiased estimate criterion function and using eigenproperties of cyclic matrices. Among several possible cyclic extensions of signal matrices, a cyclic extension leading to adaptive algorithms similar to the classical frequency domain adaptive filter is used. Since the exemplary algorithms incorporate knowledge of measurement noise, they perform better in noisy environments as compared to conventional approaches. The superior performance of the disclosed algorithms in noisy environments has been demonstrated by way of computer simulation.

Note that the exemplary algorithms of the invention differ from prior attempts to utilize a best linear unbiased estimate. For example, in the above cited paper by K.C. Ho, "A Minimum Misadjustment Adaptive FIR Filter," *IEEE Trans. Signal Processing*, Vol. 44, pp. 577-585, March, 1996 and in A.C. Orgren, S. Dasgupta, C.E. Rohrs and N.R. Malik, "Noise Cancellation with Improved Residuals," *IEEE Trans. Signal Processing*, Vol. 39, pp. 2629-2639, Dec. 1991, the measurement noise is assumed to be an autoregressive process and a whitening filter is applied to the residual signal before it is fed back to the coefficient updates. In the paper by Orgren, Dasgupta, Rohrs and Malik, it is assumed that the proper whitening filter is known *a priori,* while an adaptive whitening is considered in the paper by Ho.

By way of contrast, the algorithms of the invention do not assume any particular model for the measurement noise (although it is assumed that a correlation matrix of reasonable size describes the noise properties adequately). Additionally, the invention provides frequency domain algorithms and minimizes a variant of the best linear unbiased estimate criterion which is known for its good numerical properties. Further, according to the invention, the noise correlation is estimated during natural pauses in speech.

Those skilled in the art will appreciate that the present invention is not limited to the specific exemplary embodiments which have been described herein for purposes of illustration. The scope of the invention, therefore, is defined by the claims which are appended hereto, rather than the foregoing description, and all equivalents which are consistent with the meaning of the claims are intended to be embraced therein.

**WE CLAIM**:

1. An echo cancelation device configured to suppress an echo component of an observed signal, the echo component resulting from coupling of an echo source signal through an echo path, said echo cancelation device comprising:

an adaptive filter configured to approximate said echo path and to thereby provide an estimate of said echo component,

wherein an adaptive algorithm of said adaptive filter incorporates a measurement of a noise component of said observed signal.

2. The echo cancelation device of claim 1, wherein said measurement of said noise component of said observed signal is an *a priori* measurement made in an environment in which said echo cancelation device is designed to operate.

3. The echo cancelation device of claim 1, wherein said observed signal includes at least one intermittent speech component, and wherein measurements of said noise component of said observed signal are made in real time during pauses in said at least one speech component of said observed signal.

4. The echo cancelation device of claim 3, wherein said pauses in said at least one speech component of said observed signal are determined based upon a measurement of power in a block of samples of said echo source signal.

5. The echo cancelation device of claim 1, wherein said echo estimate is subtracted from said observed signal to provide an echo-canceled error signal, and wherein said measurement of a noise component of said observed signal is based upon an estimate of a power spectrum of said error signal.

6. The echo cancelation device of claim 1, wherein an estimated impulse response of said adaptive filter is updated when a measurement of power in a block of samples of said echo source signal is above a threshold, and wherein said estimated impulse response of said adaptive filter is not updated otherwise.

7. The echo cancelation device of claim 1, wherein said echo estimate is subtracted from said observed signal to provide an echo-canceled error signal, and wherein an estimated impulse response of said adaptive filter is computed based upon an estimate of a power spectrum of said error signal and an estimate of a power spectrum of said echo source signal.

8. The echo cancelation device of claim 7, wherein said estimated impulse response of said adaptive filter is computed based upon a weighted sum of said error signal power spectrum estimate and said echo source power spectrum estimate.

9. The echo cancelation device of claim 8, wherein said weighted sum is computed by adding said echo source power spectrum estimate to a product of said error signal power spectrum estimate and a noise matching parameter.

10. The echo cancelation device of claim 9, wherein said noise matching parameter is adjusted dynamically based upon samples of said error signal and samples of said echo source signal.

11. A method for canceling an echo component of an observed signal, the echo component resulting from coupling of an echo source signal through an echo path, said method comprising the steps of:

sampling said echo source signal to provide a block of source signal samples;

computing a discrete Fourier transform of said block of source signal samples to provide a frequency domain representation of said source signal;

multiplying said frequency domain representation of said source signal by a frequency domain estimate of an impulse response of said echo path to provide a frequency domain echo estimate;

computing an inverse discrete Fourier transform of said frequency domain echo estimate to provide a time domain echo estimate;

subtracting said time domain echo estimate from said observed signal to provide an echo canceled error signal;

computing a discrete Fourier transform of a block of samples of said error signal to provide a frequency domain representation of said error signal; and

adapting said frequency domain estimate of the impulse response of said echo path based upon said frequency domain representation of said error signal and based upon said frequency domain representation of said source signal,

wherein said step of adapting said frequency domain estimate of said impulse response incorporates a measurement of a noise component of said observed signal.

12. The method of claim 11, wherein said step of adapting said frequency domain estimate of said impulse response comprises the steps of:

updating said frequency domain estimate of said impulse response when a measurement of power in said block of source signal samples is above a threshold; and

updating an estimate of a noise power spectrum of said error signal otherwise.

13. The method of claim 12, wherein said step of updating said frequency domain estimate of said impulse response comprises the steps of:

updating an estimate of a signal power spectrum of said source signal;

computing a weighted sum of said noise power spectrum estimate and said signal power spectrum estimate; and

computing an update of said frequency domain estimate based upon said weighted sum of said noise power spectrum estimate and said signal power spectrum estimate.

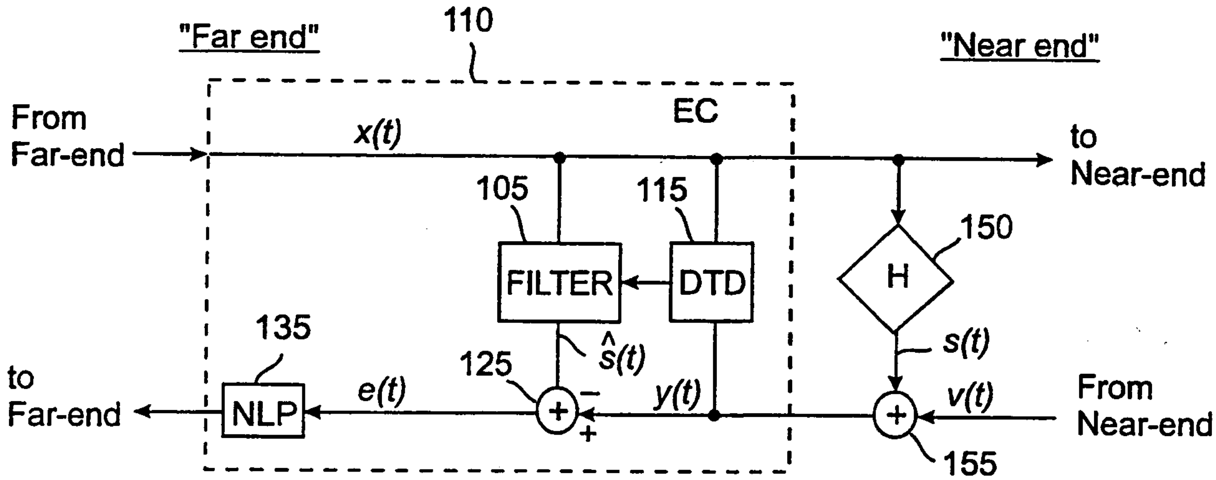
14. The method of claim 13, comprising the step of:

adding said signal power spectrum estimate to a product of said noise power spectrum estimate and a noise matching parameter to provide said weighted sum.

15. The method of claim 14, comprising the step of:

dynamically adjusting said noise matching parameter based upon samples of said echo source signal and said echo canceled error signal.

Fig. 1


"Far end"
110
"Near end"
From
Far-end
x(t)
EC
to
Near-end
105
115
150
FILTER
DTD
H
135
ŝ(t)
s(t)
125
to
Far-end
NLP
e(t)
y(t)
v(t)
From
Near-end
155

100

*Fig. 2*

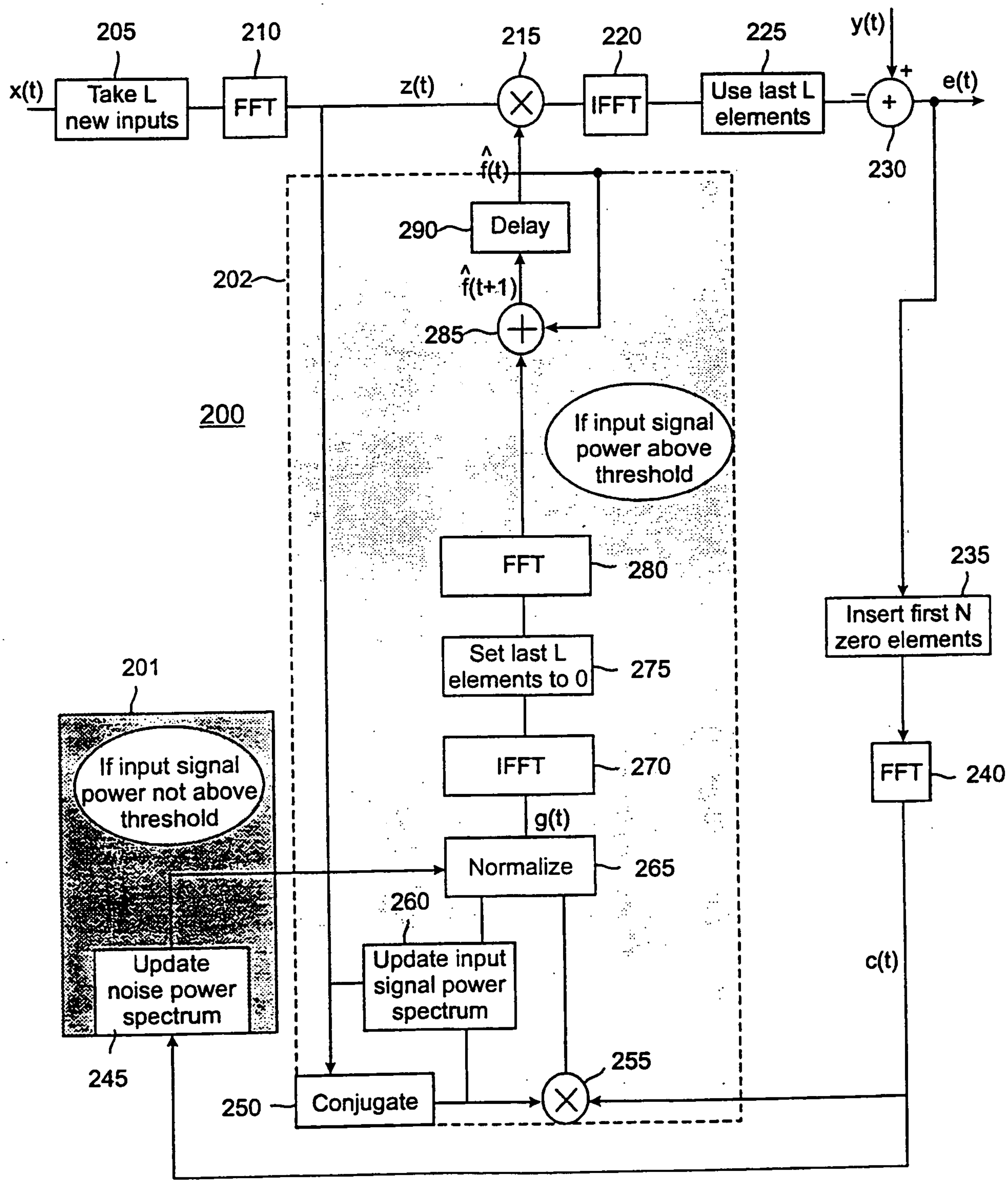

205
210
215
220
225
y(t)
x(t)
Take L new inputs
FFT
z(t)
IFFT
Use last L elements
+
–
e(t)
230
f̂(t)
290
Delay
202
f̂(t+1)
285
200
If input signal power above threshold
FFT
280
235
Insert first N zero elements
Set last L elements to 0
275
201
IFFT
270
FFT
240
If input signal power not above threshold
g(t)
Normalize
265
260
Update noise power spectrum
Update input signal power spectrum
c(t)
245
255
250
Conjugate

Fig. 3A

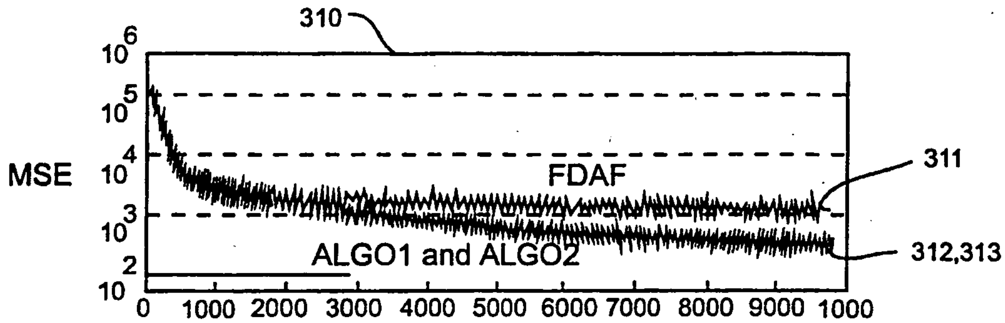

310
MSE
$10^6$
$10^5$
$10^4$
$10^3$
$10^2$
FDAF
ALGO1 and ALGO2
311
312,313
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 1000

Fig. 3B

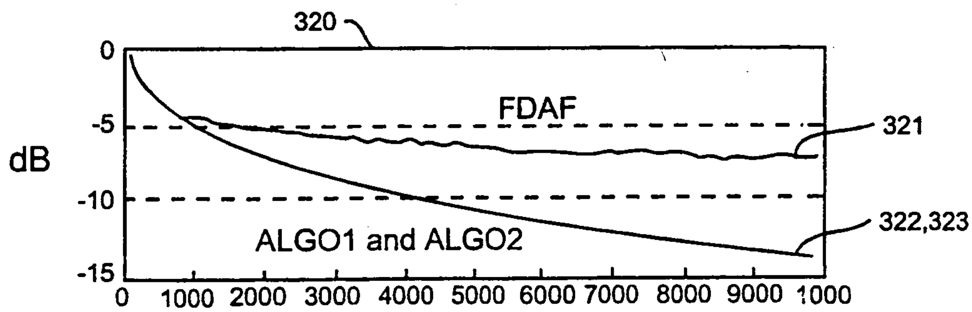

320
dB
0
-5
-10
-15
FDAF
ALGO1 and ALGO2
321
322,323
0 1000 2000 3000 4000 5000 6000 7000 8000 9000 1000

*Fig. 4*

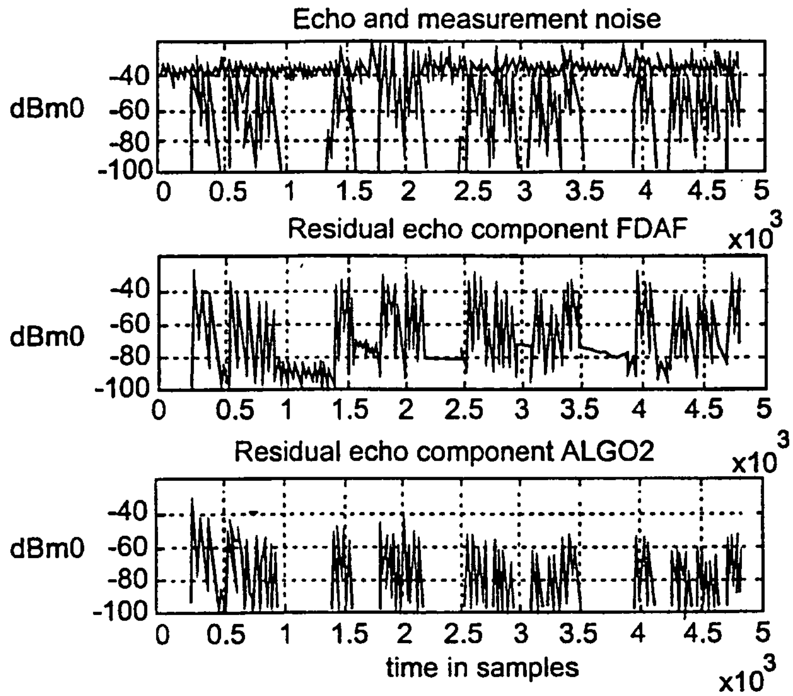

Echo and measurement noise
dBm0
-40
-60
-80
-100
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5
x10^3
Residual echo component FDAF
dBm0
-40
-60
-80
-100
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5
x10^3
Residual echo component ALGO2
dBm0
-40
-60
-80
-100
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5
time in samples
x10^3

*Fig. 5*

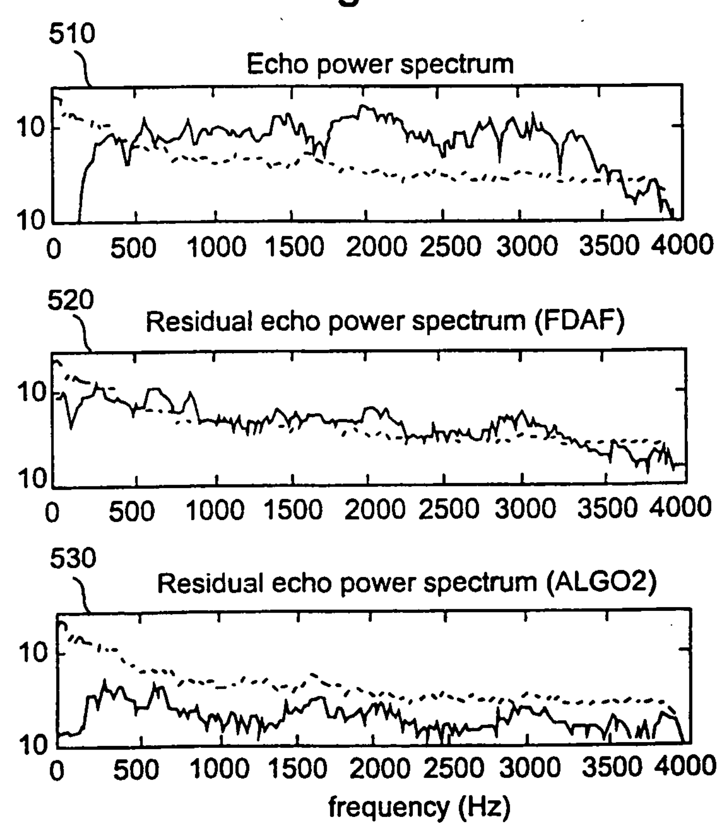

510
Echo power spectrum
10
10
0 500 1000 1500 2000 2500 3000 3500 4000
520
Residual echo power spectrum (FDAF)
10
10
0 500 1000 1500 2000 2500 3000 3500 4000
530
Residual echo power spectrum (ALGO2)
10
10
0 500 1000 1500 2000 2500 3000 3500 4000
frequency (Hz)

Fig. 6

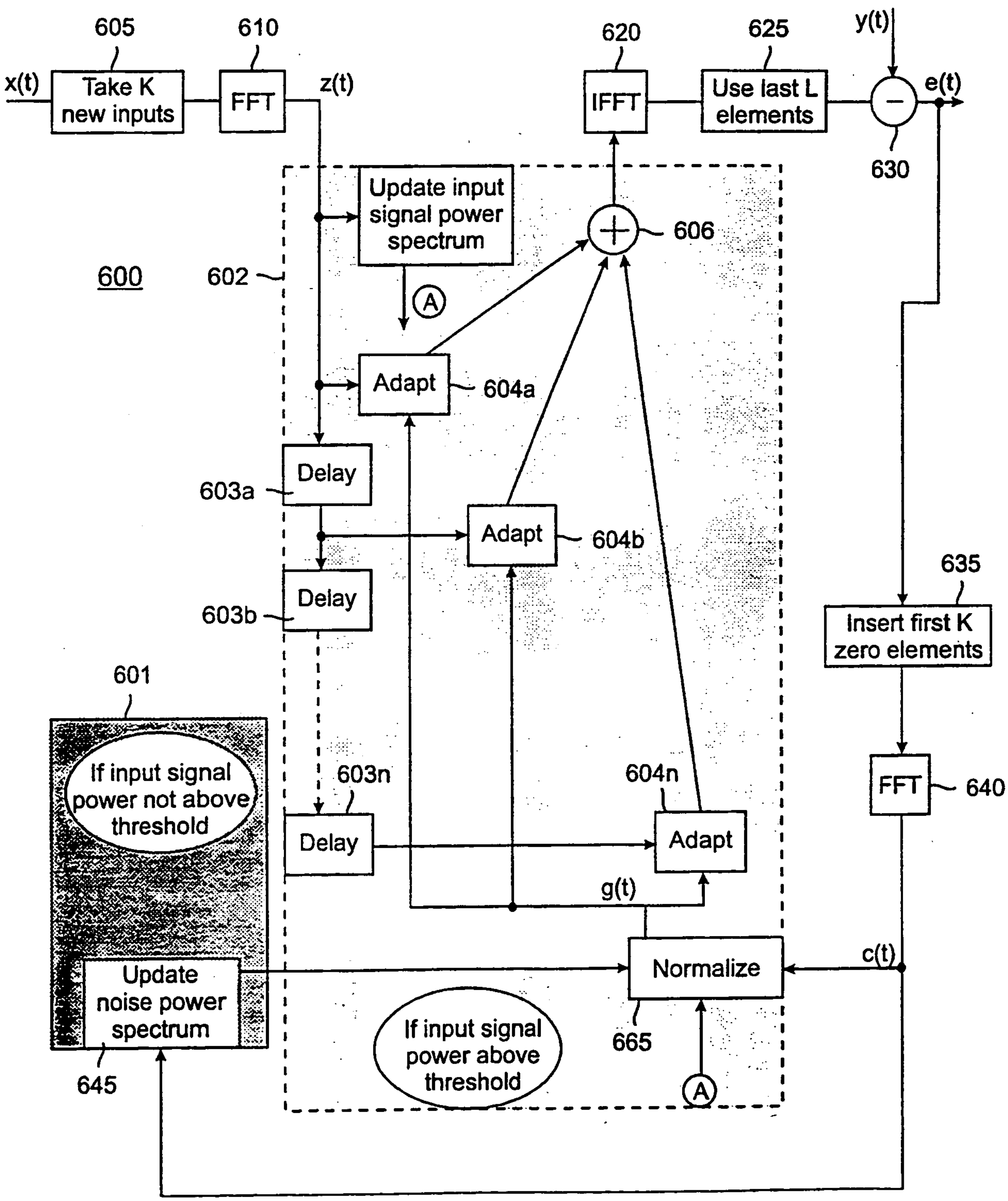
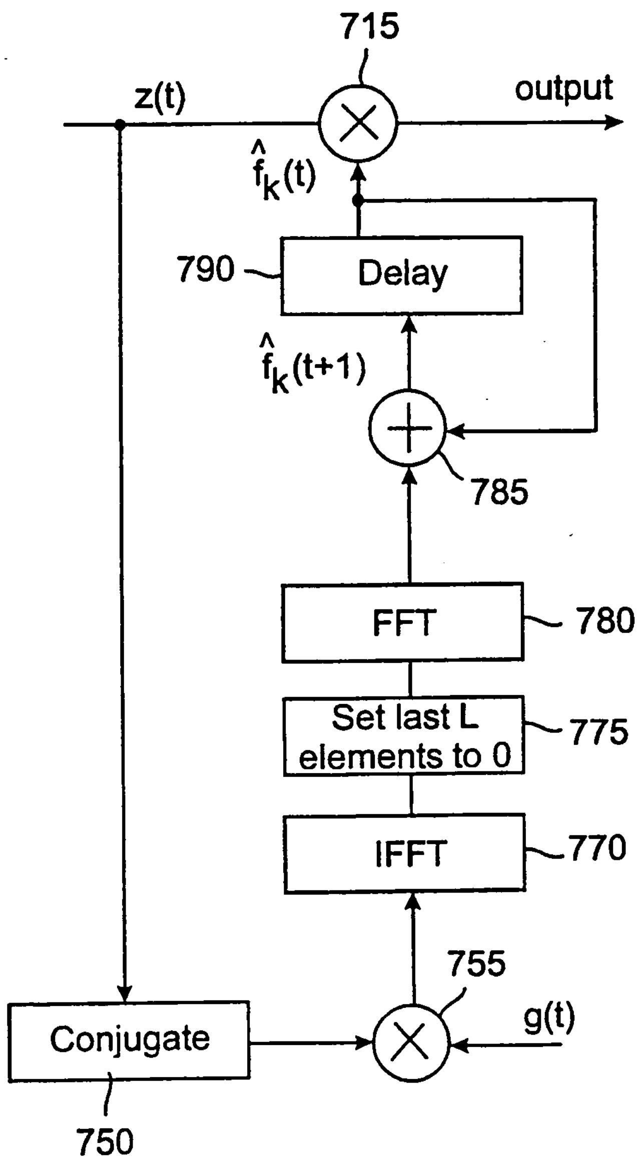

605
610
620
625
y(t)
x(t)
Take K new inputs
FFT
z(t)
IFFT
Use last L elements
e(t)
630
Update input signal power spectrum
606
600
602
A
Adapt
604a
Delay
603a
Adapt
604b
Delay
603b
635
Insert first K zero elements
601
If input signal power not above threshold
603n
604n
FFT
640
Delay
Adapt
g(t)
c(t)
Update noise power spectrum
Normalize
If input signal power above threshold
665
645
A

Fig. 7


715
z(t)
output
$\hat{f}_k(t)$
790
Delay
$\hat{f}_k(t+1)$
785
FFT
780
Set last L elements to 0
775
IFFT
770
755
g(t)
Conjugate
750

700

# INTERNATIONAL SEARCH REPORT

International Application No
PCT/SE 99/00135

**A. CLASSIFICATION OF SUBJECT MATTER**
IPC 6 H04B3/23

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 H04B H04M

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | EP 0 698 986 A (SEL ALCATEL AG) 28 February 1996<br>see claim 5<br>see abstract<br>see page 6, line 28 - line 42 | 1 |
| A | EP 0 721 252 A (AT & T CORP) 10 July 1996<br>see page 2, line 9 - line 14<br>see page 3, line 5 - line 24; figure 1 | 11 |
| A | PATENT ABSTRACTS OF JAPAN<br>vol. 097, no. 009, 30 September 1997<br>& JP 09 113350 A (NIPPON TELEGR &TELEPH CORP <NTT>), 2 May 1997<br>see abstract | 1-5,11 |

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 28 May 1999 | 07/06/1999 |

| Name and mailing address of the ISA | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,<br>Fax: (+31-70) 340-3016 | Bossen, M |

1

International Application No
PCT/SE 99/00135

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication,where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | GAY S L, TAVATHIA S: "The fast affine projection algorithm"<br>1995 INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL,<br>vol. 5, 9 - 12 May 1995, pages 3023-3026, XP002064827<br>NEW YORK, USA<br>cited in the application<br>see paragraph 2<br>----- | |

1

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/SE 99/00135

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| EP 0698986 A | 28-02-1996 | DE 4430189 A | 29-02-1996 |
| | | US 5570423 A | 29-10-1996 |
| EP 0721252 A | 10-07-1996 | US 5566167 A | 15-10-1996 |
| | | CA 2164026 A | 05-07-1996 |
| | | JP 8251085 A | 27-09-1996 |